DOCUMENT RESUME

ED 185 077 TM 800 068

AUTHOR Barnett, Nancy
TITLE The Analysis of Technical Validity and Reliability in Bilingual Language Assessment Instruments: The Language Assignment Umpire (L.A.U.) Language Dominance Test.
PUB DATE [79]
NOTE 14p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Bilingual Education; *Bilingual Students; Elementary Education; *Language Dominance; Language Fluency; *Language Tests; *Spanish Speaking; Test Bias; Test Interpretation; Test Reliability; Test Reviews; *Test Validity; *Verbal Tests
IDENTIFIERS Language Assignment Umpire

ABSTRACT

Techniques for assessing test validity and reliability are applied to an analysis of an unpublished test, in order to familiarize test users in local bilingual programs with the technical evidence that should be available for instruments of potential use in placing limited English-speaking students. The instrument studied, the Language Assignment Umpire (L.A.U.), is designed to identify language dominance by means of four oral tasks of sentence memory, synonyms, antonyms, and digit-reversal. The validity of the L.A.U. is considered in determining language proficiency as well as language dominance. Lexical difficulty and counts of phonemes, syllables, words, and morphemes are examined for a linguistic analysis of the L.A.U. The sentence memory task is examined for the syntactic complexity of its items. Statistical analyses are reported for a variety of correlations at both the intra-test and external criteria levels. A brief discussion is included of the results of a study in which the L.A.U. and other language data were used to determine the effectiveness of the Rochester, New York bilingual program. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Nancy Barnett
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED185077
TM800 068

The Analysis of Technical Validity and Reliability in Bilingual Language Assessment Instruments: The Language Assignment Umpire (L.A.U.) Language Dominance Test

Nancy Barnett
University of Rochester

A major concern in the current issue of the identification of bilingual education needs is that of the technical quality of evaluative instruments that are being used in the assessment of bilingual language proficiency and dominance. Determining the quality of such instruments is crucial, especially in that they are often relied upon as both entrance and exit criteria for bilingual and ESL programs. The following analysis of one such instrument may serve as an example of some of the types of evidence of test reliability and validity that should be looked for by test users when selecting appropriate instruments.

The current literature in language evaluation verifies many of the "abuses, misuses, and malpractices" identified by Pena and Bernal (1978). These include, to mention a few, invalid practices in test translation, test importation, inappropriate addition of points, little or no reporting of technical data, etc. In addition to these we see at times the use of valid measures in age groups for which they are not valid, unrealistic training or equipment required for proper administration of the instrument, test reliability data presented as evidence of overall test validity, or excellent reporting of technical data in a statistical form that would be incomprehensible to most individuals who might be in positions to read it or interpret it. In short, this author suggests that an unnecessary "communication gap" exists within the spectrum of those involved in bilingual language assessment. This paper addresses primarily the communication gap between test developers and test users. The research cited in this paper is reported for the purpose of identifying types of evidence that should exist for language assessment instruments. The implication is not that all local test users should conduct independent evaluations of instruments before purchasing them, but rather that users be sure to purchase only instruments that have been subject to several types of reliability and validity checks.

For this author's experience in test evaluation, as well as for part of a needed experiment in the bilingual program of the Rochester city schools, an instrument was selected for analysis that was as new as possible. This was done to avoid influence of other assessments that might have been done of an older instrument. The author also chose to work with an instrument that made use of more than one testing technique, and that did not rely upon a lot of time for administration or expensive materials. The test evaluated was the Language Assignment Umpire, which was designed by Bernard Cohen in 1976.[1] At the time of the study (Spring 1978), the test was in its field-testing phase in several areas of the country.

For this study, the L.A.U. was administered to 126 students (64 boys, 62 girls; 61 born in the United States, 61 born in Puerto Rico, 4 born elsewhere). The table below specifies the grade and program categories of the students tested. Three bilingual test administrators participated in the collection of data, all of whom were experienced in working with these age groups and were trained thoroughly in the administration of the L.A.U.

STUDENTS ADMINISTERED THE L.A.U. IN THE SPRING OF 1978

Group	Grade					School total
	4th	5th	6th	7th	8th	
A (bilingual program)	19	17	18	--	--	54
B (experimental; bilingual program for five years)	--	--	--	22	15	37
C (control; Spanish-speaking students with five years in traditional program)	--	--	--	21	14	35
	Total number of students tested:					126

(Note: For the comparison phase of research, a few students in Groups B and C were omitted because they did not meet the criterion of living in the United States for at least five years. With those students omitted, there were 19 seventh graders and 13 eighth graders in each group.)

Description of the L.A.U.

The L.A.U. author's objective is to measure a child's ability to receive language, utilize the language for cognition and produce language. Therefore, it does not contain separate sections for information regarding separate linguistic components of language, i.e., phonology, morphology, syntax, and semantics, or separate sections for the four skills areas of listening, speaking, reading, and writing. Rather it is composed of four types of verbal tasks that, in order to be carried out, depend on an overall receptive and active knowledge of the languages being tested. Since this original study was completed, the L.A.U. has been revised. Therefore, specific test items or results that have since been outdated will not be emphasized. Instead the type of analysis that was made of the instrument will be outlined.

The L.A.U. contains four parts, all of which are administered first in Spanish and then in English. Since this research involved Spanish/English bilingual students, the non-English section will here be referred to as the Spanish section. Most descriptions and general comments, however, are applicable as well to other language forms of the test. The test is presented orally and individually, with the student receiving all cues without any written or visual stimuli. Under most conditions, the test is completed within six to eight minutes.

L.A.U. Part I is a sentence memory task, in which the student is asked to repeat the phrase or sentence that the administrator reads only once. The sentences are presumably ordered in increasing length and difficulty. After three consecutive mistakes, the administrator stops and moves on to the following section. Sentences are not scored as incorrect if there is a pronunciation or intonation difference between what the administrator states and what the student repeats.[2] Neither is a response scored as incorrect if there is a syntactic change produced by the student that does not affect overall understanding of the language and concepts involved.[3]

Part II contains 28 lexical items for which the student provides synonyms. One word is given, and the student is asked to give a word that means the same thing. Correct responses are provided in the test manual for the administrator, who stops after two consecutive mistakes.

Part III of the L.A.U. is a section in which the student is asked to provide an antonym for the words given. There is again, in this section as well as in the first two sections, some flexibility of syntactic form.[4] In this section, again, the administrator stops questioning and moves on to the following section after the student has made two consecutive mistakes.

Part IV contains a digit reversal task. The student hears a combination of two, three, four, or five numbers. He hears this combination only once. As an example, we cite "3-8-2." The student then reverses the numbers, stating, in this case, "2-8-3." These items are arranged in order of difficulty, with no single digit appearing more than once in any combination of numbers. It is terminated after two consecutive erroneous responses.

Scoring of the L.A.U. is done by adding the number of correct responses per section. The sections for each language are then added, and a total score derived for each language.

Reliability

The first check of the L.A.U.'s reliability involved the inter-rater variability in scoring. A group of 37 students was taped by one rater during the administration of the test. Later, the other two raters involved in the study each listened to the tapes and rescored each of the 37 students. Inter-rater correlations were then determined for separate test parts as well as for total scores.

The second type of reliability to be determined was test/retest reliability of the L.A.U. A group of 23 students was retested after a lapse of six months. Because of the individual nature of language improvement, scores were not expected to correlate perfectly. However, a pattern of general increase in score was looked for.

Alternate form reliability was not determined for the L.A.U. because the alternate form, which is now available, was not available at the time of the original testing.

A split-half method of checking internal consistency of the L.A.U. was considered. This method is usually done systematically, e.g., odd-numbered items in one half and even-numbered items in the other half. Because of the L.A.U. procedure of terminating test parts after the student has answered two items incorrectly, this method of reliability analysis was considered to be inappropriate.

The results of the L.A.U. reliability tests indicated that there is high reliability in the instrument.[5] The slight variability in scoring creates no significant difference in comparing scores of one rater to those of another. However, since there were two raters involved in a comparison study which was a separate part of the research, the difference in rating was exactly determined, and the scores of one rater were adjusted to account for the slight variability.[6]

Validity

The validity of the L.A.U. was examined by a number of procedures. The three basic classes of validity as defined by the American Psychological Association, i.e., content, construct, and concurrent validity, would be determined by statistical and linguistic analyses.

The first of these, content validity, being primarily rational, is usually determined by the opinions of experts as to the extent that the test is a reasonable sampling of the domain being evaluated. This implies judgment of how well the test represents the domain as defined by the author of the test. Because this work was done independently, on a small scale, and not as a formal validation of the instrument, these common practices were altered a bit. First of all, the rationale and appropriateness of the sampling were evaluated without consultation with a team of experts. Secondly, each L.A.U. part was evaluated for its possible validity in measuring proficiency as well as the author's objective of identifying language dominance.

The following six issues were examined for each of the test parts:

1. The linguistic components sampled in the task.
2. The type of cognition involved in the task.
3. The objectivity in scoring of the part.
4. The appropriateness of item selection for the population tested.
5. The increase in difficulty among items.
6. The comparability of the English and Spanish sections of the part.

Part I, since it is made up of sentences rather than isolated lexical items, was analyzed in greater detail. The first four items are important to be determined in any instrument. In terms of the L.A.U., items 5 and 6 are crucial. Since the author's objective is to discriminate between various student levels by terminating the section after the student has reached three items out of his range of ability, a very gradual increase is necessary to be able to discriminate between students

of similar abilities. The comparability of difficulty of the English and Spanish sections is also crucial, especially if language totals are compared for interpretation of results.

In considering issues 5 and 6, in the sentence repetition task, item counts were taken for numbers of words, phonemes, morphemes, and syllables in each item. The examination of word and morpheme counts[7] is of interest to determine whether the progressions from sentence 1 to sentence 17 in Spanish, and from sentence 1 to sentence 17 in English, are of gradually increasing length. The task of sentence repetition demands short-term memory, which is a function of, among other things, the length of the utterance to be repeated. All other things being equal, it is assumed that the shorter of two sentences will be more easily retained for subsequent repetition. Since all other things are never equal, and since syntactic complexity and lexical familiarity appear to be important factors in short-term memory, the relative lengths of sentences must be viewed as but one contributory factor in the relative difficulty of utterances. The examination of syllable[8] and phoneme counts is also of interest in considering the progression of length within the Spanish section and the English section.

In comparing the Spanish and English sections with one another, the word and morpheme counts are of little interest, because of the synthetic nature of Spanish and the analytic nature of English. The syllable and phoneme counts, however, are of interest. The syllable count is not a completely accurate means of comparing utterance length in Spanish and English. This is due to the much higher incidence in English of long syllables (CVC, CCVC, CVCC, etc.) than in Spanish, which has a higher percentage of CV syllables (Delattre 1965:41). Combined with the phoneme count, however, it serves to give a fairly good picture of relative utterance length in Spanish and English.[9]

In addition, for all parts other than the digit reversal task, the frequency of usage of the various lexical items was considered. Word frequency lists are typically based on adult samples of written language. Adult vocabularies can contain lexical items that vary considerably from the more commonly used lexical items by children. However, since there are no lexical frequency lists available in English and Spanish for the ages involved in this research, the adult lists were used as the only criterion available.[10]

Although all of these linguistic components play some role in the relative difficulty of items, the factor of greater significance in the case of Part I is that of syntax. In an analysis of this factor, each L.A.U. item was examined for the surface complexity of its syntax and variety of verb tenses. Although it is possible to analyze the syntactic structure of these sentences in the greatest detail of their deep structure, rules, and transformations required for their production, this type of complex analysis was unnecessary to satisfy the objectives of this research.

The following table illustrates the type of syntactic description done for each Part I item:

L.A.U. PART I. SENTENCE MEMORY

Item	Tense	Syntax
	Present	NP have NP_{pl}
	Present	NP V $NP_{S\ complement}$(V NP)
	Present	N_{pl} aux-V Adv-P_{loc} Adv-P_{temp}
	Present	NP be Adv_{temp} Adj_{S}(NP_{pl} do-ing NP_{pl})
	Present	N aux-V $NP_{S\ complement}$(V NP Adv-$P[PP]_{loc}$)
	Pres. progr.	NP_{pl} aux-V Adv-P_{loc} Adv-$P[PP]_{loc}$)
	Pres. progr.	NP_{pl} and $N_{pl}{}_{S\ relative}$ (NP V Adv-$P[PP]_{temp}$)
	Present	N_{pl} aux-V NP Adv-$P[PP]_{temp}$ that $S_{subordinate\ (adverbial)}$(NP V PP)

The English and Spanish sections were found to be roughly comparable in the occurrence of auxiliaries, embeddings, and modifiers. The Spanish section at the time, however, contained many more verb tenses than did the English section.

The reader is reminded that in addressing issues 5 and 6 in Part I, no single count of any one of these linguistic factors can be relied upon solely as evidence of item difficulty, or of the comparability of the English and Spanish sections. However, it could be said that an item was inappropriately introduced if it showed a sudden increase of difficulty in several of these factors simultaneously.

As previously mentioned, the L.A.U. procedure of terminating testing of parts after a certain number of student errors calls for very gradual increase of difficulty. If one item is inappropriately placed, no great problem arises. However, if two inappropriately difficult items (three items in Part I) occur consecutively, then very often a wall is created that few students surpass. This is undesirable, in that such a "wall effect" terminates the section for students who may not be similar at all in proficiency of the particular language. If such a wall occurs in early items, this is especially serious. For comparability purposes, walls at different points, i.e., a Part I English wall occurring in items 10, 11, and 12 and a Spanish wall occurring on items 5, 6, and 7 would be especially serious, in that they would create radically different scores for students who are actually equally proficient in both languages. The L.A.U. method of totaling the number of responses rather than assigning the last item answered correctly as the score is effective in minimizing the effect of such walls.

To determine construct validity, one intra-test method carried out was that of correlating test parts to each other. Of the 126 students who were administered the L.A.U., 18 were known to be "balanced" bilinguals. This was determined if both the student and all five of his or her teachers agreed on the student's bilinguality. Although the number is small, the English and Spanish parts could be statistically correlated. If the English and the Spanish sections were in fact comparable in difficulty, then high correlations and similar ranges of responses would be expected. There were high correlations in this case, although they were misleading. The English total scores were in the range of 7-61 correct responses, while the Spanish total scores were in the range of 7-47. The high correlations, indicating a pattern of lower Spanish scores than English scores for students who are balanced bilinguals, demonstrated that revision would have to be done in making the Spanish section and the English section more comparable.

Criterion-related validity is concurrent if the two measures are administered at roughly the same time, and predictive if the measure being validated is correlated with scores of a measure that is administered after a time lapse. Concurrent validity of the English section of the L.A.U. was determined with data available from reading scores of the Metropolitan Achievement Test. Predictive validity was determined with the English section of the Language Assessment Battery, which was administered after a time lapse of six months.

Findings

Many of the results of these validity studies are both extensive to relate and in part outdated due to recent L.A.U. item revision. In general, however, it can be said that at the time, the L.A.U. item selection was more appropriate in its English section than in its Spanish section. The English section, as illustrated on Figure 1, was found to identify students who performed either very well or very poorly, but it tended to inflate the scores of some students who were in the mid-range of abilities. A small norming sample tentatively suggested that the L.A.U. Spanish scores were 1-10 points below the English scores for monolingual students of each language. A further finding was that the validity of the English section decreased as student age increased.

In terms of the validity of the L.A.U. parts, the following conclusions were drawn:

The sentence memory task was found to be a valid measure of English proficiency, especially for the younger (4th and 5th grade) students, correlating in the range of .750 - .871 with the external criteria of the reading section of the Metropolitan Achievement Test and the English section of the Language Assessment Battery. Although the English and Spanish sections were not comparable at the time in sentence length, lexical difficulty, and use of verb tenses, the conclusion was drawn that the technique of sentence memory can be valid for identifying language dominance and language proficiency.

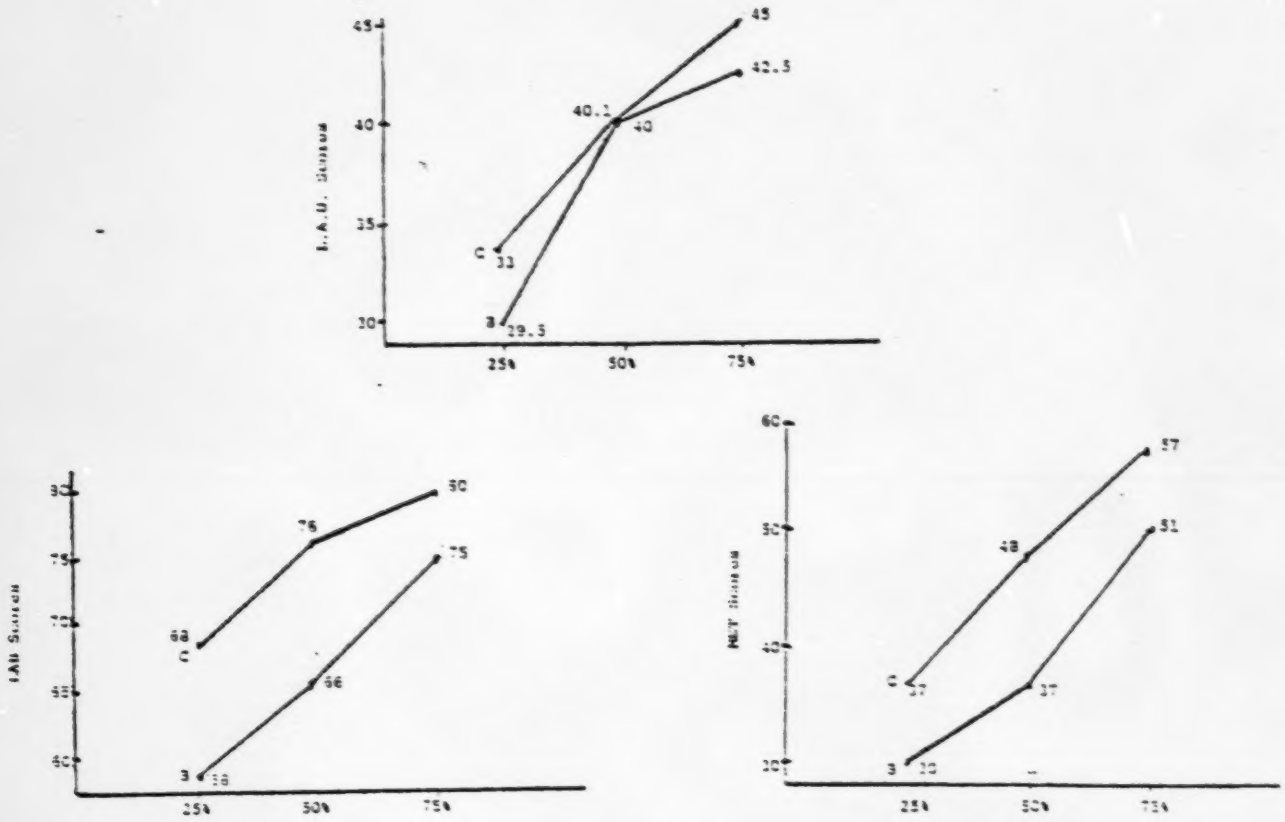

Figure 1. L.A.U. English mid-range "inflation."

The synonym task requiring both word knowledge and semantic processing did not correlate well with external criteria, with correlations suggesting that the English item selection was too easy and that the Spanish item selection was too difficult. The formation of synonyms is a skill that has been found to be difficult for bilingual children. It has been shown that bilingual children have greater flexibility in thought than monolingual children (Lambert and Tucker, 1972). Further evidence (Ben-Zeev, 1975) indicates that bilinguals are more aware of fine details in classifying words into categories than are monolinguals. This acute sense of distinction may account for the bilingual students' hesitation to respond to items with anything but an exact synonym. The acceptable L.A.U. responses to the item 'baby' ('kid,' 'child,' and 'infant') would perhaps not be acceptable to many bilingual children since 'kid' and 'child' are not perfectly synonymous with 'baby.'

These observations are not meant to suggest that synonym tasks should not be used with bilingual students. However, it is necessary that the items selected for use in this task have very closely associated synonyms and not simply related items. The 'small-tiny,' 'lindo-bonito' synonyms are more appropriate than are the 'stove-oven,' 'ver-mirar' related-item types.

The L.A.U. antonym task correlated very well with external criteria, and the conclusion was drawn that the technique can be a valid measure of determining language proficiency. At the time, the Spanish section contained more commonly used items than did the English section, and the sections are not completely comparable for the identification of language dominance.

The cognition of the production of antonyms is similar to the cognition involved in the production of synonyms. The difference is that the production of antonyms is "easier." In the administration of these tasks in the L.A.U., Vygotsky's claim (1962:88) that the child is more aware of differences than of similarities was found to be strikingly true. Many students needed several practices before understanding what was expected of them in the synonyms section, but they seemed to understand and produce opposites with much greater ease. This is due, in part, to the fact that synonyms can be represented in only one way, whereas antonyms can take the form of either contradictories, contraries, or converses.[11] Therefore, the individual is open to many more possibilities for one cue in the production of antonyms. This may suggest that the antonym task is more appropriate for younger children than is the synonym task.

The L.A.U. digit reversal task correlated poorly with the various levels of the MET and the LAB (.169 to .445). These low correlations indicate, as expected, that the task of digit reversal is only partially an indication of a student's command of language. Digit reversal, in fact, measures something more than command of language. The task is partially of the type that has been labeled as a "skill at auditory organization of verbal material," a skill which bilinguals have been found to perform better than monolinguals (Ben-Zeev, 1975). It is also related to reversal shift tasks which determine an ability to classify and reclassify data. Although it can say very little about relative language proficiency, it was found to be useful as a supplement in identifying language dominance.

There are indeed advantages and disadvantages to be found in any type of measurement instrument. Also, the careful examination to which the L.A.U. was subjected in this research would uncover methodological disadvantages in any instrument. The advantages, then, of the L.A.U. must not be overlooked.

One merit of the L.A.U. is that it incorporates four techniques into one instrument, thus avoiding the assessment of the language of children in a single way. The instrument, with some item revision, is culturally fair and can easily be transferred into several languages. It can be administered quickly, requires a minimum amount of training for its administration, and does not rely on expensive equipment or materials that could significantly increase educational costs. Once the L.A.U. has been revised, the results can be easily interpreted for educational purposes.

The L.A.U. in bilingual education

Of further interest may be the fact that the L.A.U. was used as part of a comparison study in the effectiveness of the bilingual program in Rochester. The criteria used in the selection of students in this

comparison study are listed in the table below. For this study, the L.A.U. English and Spanish scores of Spanish-speaking students who had been educated bilingually (Group B) were compared to scores of Spanish-speaking students who had been educated solely in English (Group C). Scores from the Metropolitan Achievement Test, the Language Assessment Battery, and a language classification identified by the school system were also used in the comparison. The results of the study, although simplified here, indicated that the students educated bilingually were not performing as well in English as were the Spanish-speaking students who are educated in English, and that all Spanish-speaking students in the study were performing slightly below monolingual English-speaking students of the same ages. The consistency of the lower English scores of Group B can be seen in Figure 1. The L.A.U. synonym task was the part that particularly lowered the scores of Group B students. The lower scores most probably occurred because some of the items in the synonym section called for responses that were related lexical items (e.g., "store-oven" and "baby-child") but not necessarily perfect synonyms. Although this result appears to demonstrate a negative effect of students who are educated bilingually, it may on the contrary suggest that Spanish-speaking students in traditional English instruction do not develop their bilinguality to the same extent as do students taught in both languages. Of additional interest was the superiority shown by Group B students in the digit reversal task in both English and Spanish. As expected, Group B students scored considerably higher than Group C students on all Spanish parts of the test. Quartiles of the Spanish parts demonstrated that 75% of Group C students were consistently at the level of the lowest 25% of Group B students. The results of this study pointed to 1) the need for placing greater emphasis on improving the English skills of all non-English speaking students, and especially of those educated bilingually, and 2) some linguistic and possible extra-linguistic advantages of bilingual education.

CRITERIA FOR SELECTION

Group	Grade (no. students)	Criterion
B (experimental)	7 (19) 8 (13)	Enrolled in school B (Rochester, NY) bilingual program.
		In bilingual education for the last 5 (or more) years.
		Live in the United States for at least 5 years.
		Speak Spanish since childhood.
C (control)	7 (19) 8 (13)	Enrolled in school C (Rochester, NY) traditional program.
		In traditional English instruction for last 5 (or more) years
		Live in United States for at least 5 years.
		Speak Spanish since childhood.

Summary

The reader is again reminded that the L.A.U., since the time of the reported analyses, has undergone revision. Also, the reader is reminded that the L.A.U. was in part evaluated for something for which it is not intended; i.e., its validity in determining language proficiency was considered as well as its validity in identifying language dominance. This was done, in part, out of this researcher's concern that the assumption is often made that a student is competent in the language in which he or she is classified as being dominant. The dependence of language dominance on language proficiency is an issue that should be further studied. Last, the reader is reminded that this analysis was conducted by one researcher, and that additional evaluative input is necessary for conclusive assessment.

With these factors in mind, then, several comments can be presented regarding language testing, as well as specific suggestions for users of language dominance and proficiency tests. In selecting an appropriate instrument, the following factors should be considered:

1. That the instrument has been subject to several, not just one or two, measures of reliability and validity.
2. That a team of experts, including at least one linguist, one psychologist, one statistician, and one teacher of bilingual students, has evaluated the testing rationale and item selection.
3. That statistical correlations are provided for test parts as well as totals, and that correlations are provided for all ages with which the test is to be used.
4. That comparability of the English and Spanish sections has been thoroughly studied, especially if test interpretation is done by the simple comparison of English and Spanish scores.
5. That linguistic analysis of the items is thorough, i.e., by methods other than the simple counting of words. Syntactic complexity and lexical difficulty are especially important.
6. That many acceptable responses are provided in the test manual so that administrator subjectivity is minimized as much as possible.
7. That, in addition to these suggestions, criteria such as those published by the Northwest Regional Language Laboratory be made known to test users.

It is necessary that both test users and test developers become more aware of each other's rightful concerns. With this accomplished, new instruments can be both technically valid and usable, and existing tests can be more effectively administered and understood. It is essential that test users be aware of the variety of statistical, linguistic, psychological, and other factors involved in test validation. These improvements in bilingual language assessment are necessary to insure that the linguistic abilities and needs of limited English-speaking students will be most accurately identified and these students placed in programs that are most beneficial to them.

Notes

[1]The author acknowledges the cooperation of both Bernard Cohen and the Rochester City Schools in this research.

[2]For example, features such as loss of "s" in Puerto Rican speech or variation of "r" and "l" are not scored as incorrect.

[3]For example, deletion of a single word that does not alter the meaning of the sentence. Many students repeated the sentence "My books keep on falling out of my desk during recess" without the word "on." These responses were scored as being correct.

[4]The test in its written form does not list as acceptable responses words in different classes. That is, adjectives must be matched with adjectives, nouns with nouns, etc. In the scoring, however, some responses were considered correct even if they were not of the same class as the cue word. For example, "triste" (sad) was scored as a correct response for the item "alegría" (happiness).

[5]"High" reliability refers to a correlation coefficient above .850.

[6]The procedure used for this adjustment was that of linear regression. Slopes and intercepts were obtained from original inter-rater scattergrams.

[7]For purposes of morpheme count, the complexities of Spanish verb morphology were felt to give an artificially complex picture. Therefore, for purposes of morphological counting, a verb was considered as having two morphological entities.

[8]Syllable count in Spanish was based on natural conversational speed. For example, the sentence "Canta a María" would be considered to have five syllables: can-taa-ma-rí-a.

[9]A possibly more accurate measure of utterance length, measuring recorded readings of the utterances, was not followed, since it was felt that any benefits that might be gained did not justify the time and effort involved.

[10]Lexical frequency information was obtained for both English and Spanish items in the same dictionary (Eaton 1940).

[11]Contradictories exhaust options on a scale, e.g., male-female. Contraries do not exhaust these options, e.g. large-small. Converses differ in one component, which switches in argument, e.g., parent-child. See Clark 1977:422 for details.

References

American Psychological Association. 1974. Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association.

Barnett-Garcia, Nancy, 1979. A study of the assessment of language dominance and bilingual education. Unpublished doctoral dissertation. University of Rochester.

Ben-Zeev, S. 1975. The effect of Spanish-English bilingualism in children from less privileged neighborhoods on cognitive development and cognitive strategy. Unpublished research report to the National Institute of Child Health and Human Development.

Cohen, Bernard. 1977. Language Assignment Umpire. New York: Bernard Cohen Research and Development, Inc.

DeLattre, Pierre. 1965. Comparing the Phonetic Features of English, French, German, and Spanish. London: George G. Harrap & Co., Ltd.

Eaton, Helen S. 1940. An English, French, German, Spanish Word Frequency Dictionary. New York: Dover Publications, Inc.

Lambert, W. and G. Tucker. 1972. Bilingual Education of Children--The St. Lambert Experiment. Rowley, Mass.: Newbury House.

Pena, Alban A. and Ernest M. Bernal, Jr. 1978. Malpractices in Language Assessment for Hispanic Children. Occasional Papers on Linguistics. 105-108.

Silverman, Robert J., Joslyn K. Noa, and Randall H. Russell. 1977. Oral Language Tests for Bilingual Students. Portland, Oregon: Northwest Regional Educational Laboratory.

Vygotsky, Lev Semenovich. 1962. Thought and Language. Cambridge, Mass.: MIT Press.